学科数据知识库的政策调研与分析*
——以生命科学领域为例

孙铁楠[1,3]　顾立平[1]　宋秀芳[1]　刘晶晶[1,3]　江　娴[2,3]
[1](中国科学院文献情报中心　北京　100190)
[2](中国科学院武汉文献情报中心　武汉　430071)
[3](中国科学院大学　北京　100049)

摘要: 【目的】聚焦生命科学数据知识库政策，为政策实施提供建议。【方法】通过人工阅读、筛选的方法对有明确政策声明的 38 个生命科学领域的数据知识库进行调研，主要针对数据知识库在数据提交、数据管理和数据使用等方面的政策声明。【结果】学科数据知识库的利益相关群体(数据提交者、数据管理者、数据使用者)各自的数据权益管理规范不同。【局限】仅调研 38 个生命科学领域的案例，没有对政策要素的时序变化进行分析，对政策执行细节探讨还有所欠缺。【结论】良好的学科数据知识库政策体系应当包括：数据提交政策(内容界定、格式规范、来源要求、归属说明)、数据管理声明(数据公开、数据注册、免责声明、数据版本管理)以及数据使用规范(数据访问、数据推荐引用、数据授权许可)。

关键词: 科研数据　数据知识库　政策研究　数据管理服务　生命科学
分类号: G353.1　G250.76　Q1

1　学科数据知识库政策是学科领域数据共享的推进器

科学数据是科学研究不可或缺的主要组成部分[1]。目前，国际上主要科研资助机构[2]、著名科研教育机构[3]、国际主流学术期刊[4]都纷纷制定数据政策，要求科研数据在数据知识库上进行存储与传播，并且，近期发展的数据期刊也与数据知识库相结合[5]，共同推进科研数据开放共享。

数据知识库主要分为机构数据知识库、学科数据知识库、多学科数据知识库以及特定项目数据知识库 4 类[6]。学科数据知识库是针对特定学科领域，具有更强系统化与专业化服务能力的数据知识库，受众范围广，针对程度强，倍受不同领域科研人员的青睐。学科数据库的研究重点不在于数据库本身的构建，而在于科研人员或图书馆员对数据进行提交、管理以及使用时要遵循的权利与义务，为数据提供者、数据管理者以及数据使用者之间建立起信任枢纽(Trust Relationship)。目前，学科数据知识库的政策研究主要有三方面的研究背景：

(1) 图书馆员向科研人员介绍数据管理服务(简称 DMS)、推荐数据知识库时有所依据，许多图书馆已经正在发展 DMS[7-10]；

(2) 图书馆或数据中心建设数据知识库时，完善的政策体系是必要条件，国内有许多专家正在积极从事这方面的实践工作[11-14]；

(3) 数据知识库作为数据出版与发布的主要基础设施，国内在数据开放共享上已有许多理论与实践[15-18]。

学科数据知识库的政策研究可以推进以上三方面的实践进展。

本研究聚焦生命科学领域数据知识库案例，系统

通讯作者：顾立平，ORCID：0000-0002-2284-3856，E-mail：gulp@mail.las.ac.cn。
*本文系“中国科学院文献出版领域择优支持人才”(项目编号:院1434)的研究成果之一。

地梳理 7 项子学科领域下, 利益相关者的权利与义务, 总结政策要素, 提供管理规范框架。

2 研究设计

2.1 研究思路

本研究系统地梳理和观测生命科学领域数据知识库的政策, 设计研究框架, 研究分为三个问题:

(1) 作为数据提交者, 享有哪些权利? 需要遵守哪些义务?

(2) 作为数据知识库的管理者(数据知识库本身), 权益管理有哪些?

(3) 作为数据的使用者, 享有哪些权利? 在使用数据时需要遵守哪些规范?

2.2 研究对象

根据re3data.org的记录(截至2015年7月29日), 共有1 305个数据知识库。其中生命科学(Life Sciences)领域有648个数据知识库, 572个开放共享[19]。笔者参考Scientific Data的推荐清单及分类, 选取其中网站上有明确政策的38个生命科学领域的数据知识库作为案例进行调研[20], 如表1所示:

表 1 生命科学数据知识库分类

分类	知识库
核酸序列 (Nucleic Acid Sequence)	GenBank[21]; 基因组变异数据库(DGVa)[22]; 欧洲生物信息研究所(EMBL-EBI)[23]; 欧洲核酸数据库(ENA)[24]; GO(Gene Ontology)[25]; 国际核酸序列数据库协作体(INSDC)[26]; NCBI Sequence Read Archive (SRA)[27]
蛋白质序列 (Protein Sequence)	杰克逊实验室[28]; PDBe AutoDep[29]; 通用蛋白质资源库(UniProt)[30]
分子和超分子结构 (Molecular & Supramolecular Structure)	X 射线衍射成像数据库(CXIDB)[31]; 晶体学开放式数据库(COD)[32]; 生物磁共振数据库(BMRB)[33]; Chemical Entities of Biological Interest(ChEBI)[34]; 电子显微镜数据库(EMDB)[35]; 蛋白质圆二色谱数据库(PCDDB)[36]; 全球蛋白质数据库(wwPDB)[37]
神经科学 (Neuroscience)	NeuroMorpho.Org(神经元形态数据库)[38]; 神经影像学信息工具和资源中心(NITRC)[39]; OpenfMRI[40]
组学 (Omics)	ArrayExpress 数据库[41]; GEO 数据库[42]; dbGaP[43]; 蛋白质相互作用数据库(DIPTM)[44]; PeptideAtlas[45]; 欧洲基因组表型档案网站(EGA)[46]; PRIDE 数据库[47]
生物分类及物种多样性 (Taxonomy & Species Diversity)	综合分类信息系统 (ITIS)[48]; 全球生物多样性信息设施(GBIF)[49]
生命科学学科共同体资源 (Life-science Community Resources)	鼠基因组数据库(MGI)[50]; EuPathDB[51]; FlyBase[52]; Xenbase[53]; 美国国家吸毒和艾滋病毒数据存档项目[54]; 流感研究数据库(IRD)[55]; ClinicalTrials.gov[56]; 生物通用存储库(BioGRID)[57]; 癌症和肿瘤成像存档(TCIA)[58]

3 数据提交政策

(1) 对提交的数据内容有明确界定

数据知识库通常界定收录内容, 若干生命科学领域子类别有比较详细的说明。例如, DGVa 只接受经过处理的结构变异数据, 以及与该研究有关的信息, 包括试验者/样品, 实验协议和应用分析; PDBe AutoDep 提交生物核磁共振数据; CXIDB 提交相干 X 射线衍射成像(CXI)实验数据; GEO 提交基因表达式和杂交阵列数据。由于生命领域学科的特殊性, 不同类别的生命科学数据知识库存储的数据类型不一样, 因此数据内容也就不尽相同。

(2) 对提交的数据格式应有明确规定

按照学科社群通用的软件、工具以及传输与共享的范式进行统一化规范。例如, DGVa 要求以 Excel 电子表格和/或以 TAB 作为分隔符的文本文件格式提交的数据; UniProt 鼓励用 UniProtKB/Swiss-Prot 数据的形式提交数据, 并希望提交者能提供关于所要提交数据的文字描述; BMRB 要求的格式是数据交换格式, 所有数据都以 NMR-STAR 格式存储; EuPathDB 的格式要求是所有可用序列的 FASTA 文件, 具有完整基因信息的 GFF 文件, 以及其他格式文件。

(3) 要求数据提交者提交的数据符合科学伦理

数据知识库的提交政策中往往包括一项声明, 要求数据提交者所提交的数据不能来自违反科学伦理的实验或者采集, 即: 编辑出版不会传播违背科学良知的行为产物。这意味着存储在数据知识库中的数据本身不存在伦理道德的争议。例如, GenBank 规定如果提

交的数据涉及到人类序列数据，则不能包括任何可能揭示数据提供者个人身份的数据。

(4) 需要有明确的数据归属说明

著作权包括精神权利和财产权利。精神权利是不可剥夺的权利，例如数据提供者的署名。财产权利需要数据知识库明确声明，数据知识库的管理团队和负责人也需要在数据提交政策中明确在数据提交后将会如何处理数据等。例如，ArrayExpress 规定，提交的数据由专门的生物管理团队进行管理，或者每周从 NCBI 基因表达式综合数据库系统地导入。

数据提交政策是任何类型数据知识库政策所应该有且必须有的政策基础。科研人员向数据知识库提交数据之后，更加关注的是自己的数据是否可以得到充分的保护以及合理的传播使用，这是建立数据提交者与数据知识库管理者之间相互信任桥梁的基石。因此，数据管理声明和数据使用规范是数据知识库政策研究的重点。

4 数据管理声明

4.1 鼓励数据提交者尽快公开数据

数据的公开途径与方式是数据知识库管理政策的基础，数据的使用、传播、更改等行为都是基于数据在严格且正确的公开途径下进行。数据公开的途径通常有三类:

(1) 数据提交者获得一组账号和密码，能够自己使用或者课题组内使用;

(2) 除了数据提交者外，学术同行也能使用数据，这主要是用于学术出版所需的同行评审;

(3) 在数据提交后，数据完全开放，任何人都可以使用该数据。

数据公开途径可能只有上述一种，也可能经历多种途径，这就涉及到数据公开的时间以及数据知识库的政策。例如，EMDB 的数据公开途径就是第三种情况，该数据知识库鼓励提交者尽快公开发布他们的数据，而不是搁置他们的数据。

4.2 为数据提交者分配编号 ID

发表在学术期刊上的学术论文通常标识着作者、题名、刊名、卷期号、页码等，无论发布在哪个平台上都可被识别并作为引用对象。科研数据也可能存储在不同的发布平台上，这就需要有一套认证机制来保护科研人员的权益。目前，有三类常用途径：通用的数字资源唯一标识符，如 DOI[59]等；作者唯一标识符，如 ORCID[60]等；数据库的 URNs 或数据编号(Accession No.，简称“编号 ID”)。以生命科学领域而言，数据知识库分配编号 ID 往往已在领域内形成认可和识别科研数据产生的一套约定俗成的规范，许多生命科学数据知识库有其分配编号 ID 的服务机制。例如，DGVa 在对数据存档后，会给研究对象、变体区域以及所有样本层面的变体(识别变体)分配唯一且稳定的编号 ID；wwPDB、BMRB 也均会为数据提交者提交的数据分配相应编号 ID。可见，数据知识库分配编号 ID 已成为科研人员投稿学术期刊论文、发表数据期刊论文、注释及引用自己和他人数据集的既成规范。

4.3 应说明使用风险、产权保护、使用条款等

互联网上的内容服务网站通常都有若干声明，本文所说的免责声明主要是指数据知识库在数据管理政策上所涉及的声明。本次调研的多个数据知识库都涉及相关免责声明，由于生命科学领域的科研数据内容多样、来源各异且性质复杂，大多数的数据知识库都会有相应的免责声明，可细化为三项重点:

(1) 数据使用风险的免责声明。数据知识库通常会提出：数据知识库不会出于任何目的，对软件或数据的适用性和准确性作出说明，也不会作出任何保证、明示或暗示，包括可销售性和对于特定用途的适用性，用户需自己承担风险。

(2) 数据知识产权的免责声明。数据知识库一般不负责评估原始数据的提交者对已提交的全部或部分数据要求专利、版权或其他知识产权的有效性，也没有权利将数据转移给第三方，同时，不能保证使用软件或数据不会侵犯任何第三方的专利、版权、商标或其他权利。

(3) 数据访问过程的免责声明。用户对数据库中数据的任何使用或者误用，以及对于通过任何与数据库相关的浏览器、客户端或第三方软件进行访问时出现的任何问题，数据库不承担任何责任。

此外，一般国外网站都有的使用条款(如拒绝批量下载等)和隐私条款(如对使用日志的妥善处理)等。免责声明是数据知识库管理规范中必不可少的内容，其中，数据使用风险和数据知识产权是比较特殊的免责声明，可按照实际情况进一步条款化。

4.4 要求数据提交者发布最新版本的数据

存储的科研数据通常有三种情况: 科研数据本身已经不会再做任何更动, 科研数据来自于一次性的实验或观察, 经过整理后所提交的数据已是最终版本; 数据本身是长期动态更新的, 这类数据随着时间的变化会不断增加实质性内容, 因此很难界定所谓的最终版本, 反而需要以 V1, V2, V3, …Vn 的形式标识各个时间段的不同版本; 出于不同阶段对数据发布或出版的要求以及对数据公布(参考 4.1 节)的情况, 可能会有内容完整性上的差异, 需要进行数据版本的区别和管理, 这种情况可能会与前两种情况存在重合; 另外有一种特殊情况, 或称为第 4 种情况, 数据可能被其他人或科研团队本身发现存在错误、瑕疵和遗漏等问题, 需要予以更新。在具体实施时需要注意以下 4 点问题:

(1) 数据知识库只呈现数据的最新版本, 但不删除过去所有版本的数据。INSDC 声明, 数据提交者对数据进行修改和更新时, 修正后的数据将从下一个数据知识库版本中删除, 但所有的数据都将继续通过登录号实现永久访问。有的数据库对修正后的数据文件赋予新的登录查询号, 保证用户对每个版面的持久访问, 并同时提供原始数据集和修正后的数据文件。

(2) 每次更新或修改数据版本时都需要提交说明文件。例如, ClinicalTrials.gov 规定数据提交者在对数据进行修改或重新分配数据知识库的修改版本时, 需要用完整的说明文档进行声明或申请权限。

(3) 数据知识库管理者更动数据的特殊情况和实施办法。例如, BMRB 规定数据知识库的工作人员可以修改文件以更新引用, 或升级文件以保持数据格式的一致性。但需要在修订版本文件的开头添加描述修订日期、修订作者的声明。

(4) 数据知识库的每一个新版本是补充原有数据集的, 而非替换。在修改原始数据集时, 有的数据知识库(例如 PRIDE)会为提交者提供 FTP 详细资料用来上传数据, 并将添加到原来的数据集, 而不是重新提交整个数据集。综上, 就学科数据知识库而言, 通常会令数据提交者发布最新版本的数据, 较少或多数不愿以新上传的数据集覆盖已有的数据集。

5 数据使用规范

无论是数据的提交还是数据的管理, 数据知识库的最终目的是为科研人员提供数据使用和传播服务, 这才是真正的价值所在。数据知识库出于对数据提交者权益的保护以及对数据使用者的责任, 必须制定数据使用规范, 为学术社群提供良好且健康的数据共享环境。

5.1 应明确数据访问权责

数据知识库的用户通常可以上传、浏览、下载科研数据甚至使用数据知识库提供的工具对科研数据进行分析, 但有些数据库会区分不同类型的用户, 如注册用户、高级用户和普通用户等。有些数据知识库要求科研人员注册后才能够上传或下载科研数据(作为数据贡献者之一才能作为数据使用者), 有些数据知识库要求必须成为会员(科研机构或项目团队支付或分担一部分的数据知识库维护费用), 有些数据知识库不需要用户经过注册或成为会员就可以使用。

核酸序列、蛋白质序列、分子和超分子结构、神经科学、组学和生命科学学科共同体资源类别下的数据知识库均有数据访问政策的内容:

(1) 一般情况下, 数据知识库普遍支持科研数据的开放共享和传播, 在使用政策中说明了可以自由地下载、使用或传播。

(2) 若干数据知识库还会考虑个别数据提供者在提交时出于商业合同(例如数据集中存在采购得来的授权数据)、项目资助者要求(例如有可能提到可以浏览、下载和利用数据验证科研成果, 但不允许利用数据形成新的产品)、项目保护需求(例如尚未完成或成果尚未发表)等, 对若干使用条件作出限制或者写明前提条件。

(3) 个别数据知识库允许用户以编程的方式开发数据分析流程、网站或数据视图, 以便更方便快捷地访问数据和集成公共数据, 例如 EMBL-EBI。

(4) 有些数据使用者出于检验科研成果或进行其他方面科研等的目的, 具有接触某一数据集的受测对象的需求, 而数据知识库必须保障数据提供者的相关权益, 通常采取的做法是让数据使用者提交申请进行审批。

(5) 数据知识库必须有相关声明, 在使用数据进行专利申请时, 需要获得数据提供者的许可或科研机构、科研资助机构的商业用途许可证。例如, TCIA 的相关规定。

5.2 规范数据引用格式

科研人员采集、整理、汇总、分析和提供科研数据是一种支持和促进科学发展的工作，人们利用这些数据研究新的问题，也不应忽视数据提供者的科研贡献。因此，科研数据的引用不仅对于完善科研评价体系具有一定的作用，而且对数据知识库能够长期发展也扮演着关键角色。

数据知识库的数据引用有若干种情况：要求引用数据集的标识符，可能是数字资源标识符或数据知识库分配的编号 ID；要求引用数据知识库本身的名称及其网址；要求引用数据知识库创建者所撰写的一篇论文。生命科学领域数据知识库普遍倾向引用数据知识库及其信息。例如：

(1) ITIS 的引文格式：检索[月、日、年]，来自分类学信息系统在线数据库，http://www.itis.gov;

(2) IRD 推荐的引用方式：Squires et al. (2012) 流感研究数据库：用于流感研究和监测的综合生物信息学资源。流感及其他呼吸道病毒 DOI: 10.1111/ j.1750-2659.2011.00331.x 等。

5.3 个别数据知识库需要明确数据授权许可

数据授权许可协议是数据开放共享的权利规范，能够为进一步发挥科研数据的巨大科学价值、经济价值、社会价值起到保驾护航的作用。数据授权许可证是一整套的规范，涉及到使用者被允许进行的行为活动：浏览阅读、下载存储、复制传播、内容重用、内容挖掘、信息抽取、内容汇编、重制衍生、附加许可、商业使用等。目前，数据知识库所讨论的授权许可主要着重在元数据和数据的许可证协议(License)，通常采用知识共享协议。

生命科学学科共同体资源类别下的数据知识库的政策指出：

(1) 禁止以营利为目的的传播。例如，MGI 规定事先未得到明确的书面许可，数据禁止任何用于商业用途的传播和使用。

(2) 禁止商业用途的信息抽取和利用。例如，ClinicalTrials.gov 规定用户不得从数据知识库中提取电子邮件地址用于营销或其他促销行为。

(3) 特殊信息的使用授权。例如，美国国家吸毒和艾滋病毒数据存档项目规定对于机密数据，获得机密数据任何部分的研究人员和机构有义务保护个人的机密信息，采取预防措施以保护机密数据的非授权使用。

(4) 非专属权。例如，ClinicalTrials.gov 声明用户不拥有数据知识库任何部分的任何专有权，也不能代表数据库或其任何部分。

6 结 语

(1) 政策管理规范框架

不同学科领域的学术规范各有不同，即便在生命科学领域中的子类别中也略有差异，因此学科数据知识库的政策研究重点不在“政策应该是什么”而是“政策制定需要考虑什么”。建议根据生命科学领域课题的研究特性，结合表 2 的管理规范框架，实行相关措施。

表 2 生命科学数据知识库的管理规范框架

政策规范	要素	实施建议
数据提交政策	数据内容界定	存储的数据类型不一样，数据内容也不尽相同。
	数据格式规范	依据学科社群约定成俗的规范进行制定。
	数据来源要求	符合科学伦理和学术道德。
	数据归属说明	保证数据提供者的精神权利和财产权利。
数据管理声明	数据公开	数据知识库要求数据尽可能开放共享。
	数据注册	数据知识库给数据分配编号 ID 作为标识符。
	免责声明	使用风险；知识产权；访问过程；其他。
	数据版本管理	更新或修改数据版本时都需要提交说明文件。
数据使用规范	数据访问	通常可以自由下载、使用或传播；特殊情况，另定审核办法。
	数据推荐引用	数据知识库名称、网址以及数据注册的编号 ID。
	数据授权许可	通常采用 CC/BY 协议，也强调特殊保密和非专属权。

(2) 实践推广的建议

本研究的实践意义有两方面：如果机构要建立学科数据知识库或以机构知识库存储和传播科学数据时，可以参考本文所提出的管理规范框架(表 2)；如果图书馆员要推广国内外合适的学科数据知识库给科研人员时，可以从学科分类以及政策完善性两方面进行推介。

(3) 研究局限

需要注意的是，对研究结论进行推广时还需进行其他新的验证。如果在我国制定有关的政策内容，建

议结合科研人员的需求和政策专家团队的咨询意见。另外，本研究没有对生命科学领域数据知识库政策要素的时序变化进行分析，对政策执行细节也还有所欠缺，这可能需要后续与该领域的专家进行讨论。

(4) 未来研究

在本研究的基础上，结合对科研人员的访谈调查，形成一系列实践指南，如《学科数据管理计划指南》等。

(致谢：中国科学院文献情报中心组织编译了《生命科学数据知识库案例汇编》报告，并在机构知识库(ir.las.ac.cn)上提供开放共享，为本研究提供了参考。)

参考文献：

[1] 刘荣，张娜. 共享体现科学数据价值——访中国工程院院士孙九林[J]. 科技创新与品牌，2011(7): 10-13. (Liu Rong, Zhang Na. Sharing Reflect the Value of Research Data—Visit to Chinese Academy of Engineering Sun Jiulin [J]. Science Technology Innovations and Brands, 2011(7): 10-13.)

[2] Jones S. Developments in Research Funder Data Policy [J]. International Journal of Digital Curation, 2012, 7(1): 114-125.

[3] Davidson J, Jones S, Molloy L, et al. Emerging Good Practice in Managing Research Data and Research Information within UK Universities [J]. Procedia Computer Science, 2014, 33: 215-222.

[4] Scientific Data to Complement and Promote Public Data Repositories [EB/OL]. [2015-09-07]. http://blogs.nature.com/scientificdata/2013/07/23/scientific-data-to-complement-and-promote-public-data-repositories/.

[5] 刘凤红，崔金钟，韩芳桥，等. 数据论文：大数据时代新兴学术论文出版类型探讨[J]. 中国科技期刊研究，2014, 25(12): 1451-1456. (Liu Fenghong, Cui Jinzhong, Han Fangqiao, et al. Data Paper: New Types of Academic Papers Published in Big Data [J]. Chinese Journal of Scientific and Technical Periodicals, 2014, 25(12): 1451-1456.)

[6] Pampel H, Vierkant P, Scholze F, 等. 呈现科研数据知识库: re3data.org 注册机制[J]. 顾立平译. 现代图书情报技术，2014(3): 26-34. (Pampel H, Vierkant P, Scholze F, et al. Making Research Data Repositories Visible: The re3data.org Registry [J]. Translated by Ku Liping. New Technology of Library and Information Service, 2014(3): 26-34.)

[7] 黄如花，邱春艳. 图书馆参与科学数据管理中的元数据应用实践研究[J]. 图书与情报, 2014(5): 65-69. (Huang Ruhua, Qiu Chunyan. Research on Metadata Application Practices of Library Participating in Data Curation [J]. Library & Information, 2014(5): 65-69.)

[8] 司莉，邢文明. 国外科学数据管理与共享政策调查及对我国的启示[J]. 情报资料工作，2013, 34(1): 61-66. (Si Li, Xing Wenming. Scientific Data Management and Sharing Policies in Foreign Countries: Investigation and Inspiration to Us [J]. Information and Documentation Services, 2013, 34(1): 61-66.)

[9] Borgman C L. 科研数据共享的挑战[J]. 青秀玲译. 现代图书情报技术, 2013(5): 1-20. (Borgman C L. The Conundrum of Sharing Research Data [J]. Translated by Qing Xiuling. New Technology of Library and Information Service, 2013(5): 1-20.)

[10] 顾立平. 科学数据权益分析的基本框架[J]. 图书情报知识, 2014(1): 34-51. (Ku Liping. Basic Framework of the Analysis on Research Data Rights [J]. Documentation, Information & Knowledge, 2014(1): 34-51.)

[11] 诸云强，孙九林，冯敏，等. 论地学科研信息化环境[J]. 中国科学院院刊，2013, 28(4): 501-510. (Zhu Yunqiang, Sun Jiulin, Feng Min, et al. Study on e-Science for Geosciences [J]. Bulletin of Chinese Academy of Sciences, 2013, 28(4): 501-510.)

[12] 马俊才，刘斌，吴林寰，等. 利用科研信息化手段推动微生物研究与应用[J]. 中国科学院院刊，2013, 28(4): 519-524. (Ma Juncai, Liu Bin, Wu Linhuan, et al. Promote Microbiological Research and Application by Virtue of e-Science [J]. Bulletin of Chinese Academy of Sciences, 2013, 28(4): 519-524.)

[13] 黎建辉. 中国科学数据建设与共享[EB/OL]. [2015-07-20]. http://ir.las.ac.cn/handle/12502/7444. (Li Jianhui. Scientific Data Construction and Sharing of China [EB/OL]. [2015-07-20]. http://ir.las.ac.cn/handle/12502/7444.)

[14] 刘峰，张晓林，孔丽华. 科研数据知识库研究述评[J]. 现代图书情报技术, 2014(2): 25-31. (Liu Feng, Zhang Xiaolin, Kong Lihua. Research Review on the Research Data Repositories [J]. New Technology of Library and Information Service, 2014(2): 25-31.)

[15] 诸云强，孙九林，王卷乐，等. 论地球数据科学与共享[J]. 国土资源信息化, 2015(1): 3-9. (Zhu Yunqiang, Sun Jiulin, Wang Juanle, et al. Study on Earth Data Science and Data Sharing [J]. Land and Resources Informatization, 2015(1): 3-9.)

[16] 刘闯，王晋年，吕婷婷，等. 大湄公河次区域中尺度行政区划地理信息系统数据库(GMS_AdmBnd)开发标准研究[J]. 地理信息世界，2010, 8(1): 17-26, 42. (Liu Chuang, Wang Jinnian, Lv Tingting, et al. Standard for GMS_AdmBnd An Updated Moderate Scale Administrative Boundary GIS Database of Great Mekong Subregion [J]. Geomatics World, 2010, 8(1): 17-26, 42.)

[17] 张计龙，殷沈琴，张用，等．社会科学数据的共享与服务——以复旦大学社会科学数据共享平台为例[J]．大学图书馆学报，2015，33(1)：74-79．(Zhang Jilong, Yin Shenqin, Zhang Yong, et al. Social Scientific Data Sharing and Serving——An Example of Fudan University Social Scientific Data Platform [J]. Journal of Academic Libraries, 2015, 33(1): 74-79.)
[18] 屈宝强，吴家喜，赵伟，等．地方科技文献共享平台服务研究[J]．国家图书馆学刊，2012，21(1)：68-72．(Qu Baoqiang, Wu Jiaxi, Zhao Wei, et al. Analysis of Information Services of Local Scientific and Technical Literature Sharing Platform Based on Website Information [J]. Journal of the National Library of China, 2012, 21(1): 68-72.)
[19] re3data [EB/OL]. [2015-07-29]. http://www.re3data.org.
[20] Scientific Data [EB/OL]. [2015-07-29]. http://www.nature.com/sdata/data-policies/repositories.
[21] GenBank [EB/OL]. [2015-07-20]. http://www.ncbi.nlm.nih.gov/genbank/.
[22] DGVa [EB/OL]. [2015-07-20]. https://www.ebi.ac.uk/dgva.
[23] EMBL-EBI [EB/OL]. [2015-07-20]. https://www.ebi.ac.uk/.
[24] ENA [EB/OL]. [2015-07-20]. https://www.ebi.ac.uk/ena.
[25] Gene Ontology [EB/OL]. [2015-07-20]. http://geneontology.org/.
[26] INSDC [EB/OL]. [2015-07-20]. http://www.insdc.org/.
[27] NCBI Sequence Read Archive [EB/OL]. [2015-07-20]. http://www.ncbi.nlm.nih.gov/Traces/sra/.
[28] Jackson Laboratory [EB/OL]. [2015-07-20]. http://www.jax.org/.
[29] PDBe AutoDep [EB/OL]. [2015-07-20]. http://www.ebi.ac.uk/pdbe/.
[30] UniProt [EB/OL]. [2015-07-20]. http://www.uniprot.org/.
[31] CXIDB [EB/OL]. [2015-07-20]. http://www.cxidb.org/.
[32] COD [EB/OL]. [2015-07-20]. http://www.crystallography.net/.
[33] BMRB [EB/OL]. [2015-07-20]. http://www.bmrb.wisc.edu/.
[34] ChEBI [EB/OL]. [2015-07-20]. https://www.ebi.ac.uk/chebi/init.do.
[35] EMDataBank [EB/OL]. [2015-07-20]. http://www.emdatabank.org/index.html.
[36] PCDDB [EB/OL]. [2015-07-20]. http://pcddb.cryst.bbk.ac.uk/home.php.
[37] wwPDB [EB/OL]. [2015-07-20]. http://www.wwpdb.org/.
[38] NeuroMorpho.Org [EB/OL]. [2015-07-20]. http://neuromorpho.org/neuroMorpho/index.jsp.
[39] NITRC [EB/OL]. [2015-07-20]. http://www.nitrc.org/.
[40] OpenfMRI [EB/OL]. [2015-07-20]. https://openfmri.org/.
[41] ArrayExpress [EB/OL]. [2015-07-20]. http://www.ebi.ac.uk/arrayexpress/.
[42] GEO [EB/OL]. [2015-07-20]. http://www.ncbi.nlm.nih.gov/geo/.
[43] dbGaP[EB/OL]. [2015-07-20].http://www.ncbi.nlm.nih.gov/gap/.
[44] DIPTM [EB/OL]. [2015-07-20]. http://dip.doe-mbi.ucla.edu/dip/Guide.cgi.
[45] PeptideAtlas [EB/OL]. [2015-07-20]. http://www.peptideatlas.org/overview.php.
[46] EGA [EB/OL]. [2015-07-20]. https://www.ebi.ac.uk/ega/.
[47] PRIDE [EB/OL]. [2015-07-20]. https://www.ebi.ac.uk/pride/archive/.
[48] ITIS [EB/OL]. [2015-07-20]. http://www.itis.gov/.
[49] GBIF [EB/OL]. [2015-07-20].http://www.gbif.org/.
[50] MGI [EB/OL]. [2015-07-20]. http://rgd.mcw.edu/.
[51] EuPathDB [EB/OL]. [2015-07-20]. http://eupathdb.org/eupathdb/.
[52] FlyBase [EB/OL]. [2015-07-20]. http://flybase.org/.
[53] Xenbase [EB/OL]. [2015-07-20]. http://www.xenbase.org/entry/.
[54] NAHDAP [EB/OL]. [2015-07-20]. http://www.icpsr.umich.edu/icpsrweb/NAHDAP/index.jsp.
[55] IRD [EB/OL]. [2015-07-20]. http://www.fludb.org/brc/home.spg?decorator=influenza.
[56] ClinicalTrials.gov [EB/OL]. [2015-07-20]. https://clinicaltrials.gov/.
[57] BioGRID [EB/OL]. [2015-07-20]. http://thebiogrid.org/.
[58] TCIA [EB/OL]. [2015-07-20]. http://www.cancerimagingarchive.net/.
[59] 毛军，孟连生，镇锡惠，等．试论我国数字资源唯一标识符发展战略[J]．现代图书情报技术，2005(2): 1-4. (Mao Jun, Meng Liansheng, Zhen Xihui, et al. Establish a Framework of Digital Resource Unique Identifier System in China: Strategy and Economics [J]. New Technology of Library and Information Service, 2005(2): 1-4.)
[60] 刘润达，王运红．开放研究人员及贡献者唯一标识(ORCID)概述[J]．情报科学，2013，31(11)：86-90．(Liu Runda, Wang Yunhong. An Overview of Open Researcher & Contributor ID [J]. Information Science, 2013, 31(11): 86-90.)

作者贡献声明：

孙轶楠：论文主体撰写，政策细节梳理；
顾立平：研究设计，信息源提供，政策要素分析，论文终稿修订；
宋秀芳：调研对象的遴选与研究；
刘晶晶：提供 Scientific Data 数据知识库清单，参与案例收集，参与论文讨论与修改；
江娴：负责主要生命科学数据知识库的案例编译。

收稿日期: 2015-08-03
收修改稿日期: 2015-09-10

The Policy Research and Analysis of Subject Data Repository ——Cases Study of Life Sciences

Sun Yi'nan[1,3] Ku Liping[1] Song Xiufang[1] Liu Jingjing[1,3] Jiang Xian[2,3]
[1](National Science Library, Chinese Academy of Sciences, Beijing 100190, China)
[2](Wuhan Library, Chinese Academy of Sciences, Wuhan 430071, China)
[3](University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: [**Objective**] For policy implementation recommodation, this study focuses on life sciences data repositories policies. [**Methods**] By artificial reading, screening method to investigate 38 data repositories in the field of life sciences having a clear policy statements, this paper summarizes and analyzes these policy statements mainly from data repository in data submission, data management and data using. [**Results**] The stakeholder groups of Subject Data Repository (data administrator, data contributors and data users) have different data rights management specification. [**Limitations**] Just simply investigate 38 cases in the field of life sciences, without analyzing the temporal change of policy elements and lacked policy implementation details. [**Conclusions**] Good policy system of subject data repository should include: data submission policy (including content definition, format specification, source requirements and belonging instructions), data management statement (including data disclosure, data registration, disclaimer and data version management), and use of data specification (including data access, data recommended reference and data licensing).

Keywords: Research data Data repository Policy research Data management service Life sciences